Supplementary Material for

Long signal peptides of RGMa and DCBLD2 are dissectible into subdomains according to the NtraC model

Eduard Resch, Jan A. Hiss, Alexander Schreiner, Gisbert Schneider, Anna Starzinski-Powitz

in Molecular BioSystems.

Supplementary Materials and Methods

Cloning of the RGMa SEAP constructs iso1-N-RGMa, iso2-N-RGMa, iso1-SP-RGMa, iso2-SP-RGMa

Iso1-N-RGMa and iso2-N-RGMa were cloned by PCR using N-RGMa (described in materials and methods of the main document) as template and the following forward primers: F*Not*I_iso1-RGMa 5'-TTG CGG CCG CAT GGG TGG CCT GGG GCC ACG ACG GGC GGG AAC CTC GAG GGA GAG GCT AGT GGT AAC AGG-3' and F*Not*I_iso2-RGMa 5'-TTG CGG CCG CAT GGG TAT GGG GAG AGG GG-3'. Iso1-SP-RGMa and iso2-SP-RGMa were created by PCR with the same primers using SP-RGMa (described in materials and methods of the main document) as template. In all reactions the oligonucleotide *Acc65*I SEAP Myc-tagging 5'-TTG GTA CCT TAC AGA TCC TCT TCT GAG ATG AGT TTT TGT TCA CCC GGG TGC GCG GCG TCG-3' served as reverse primer. All constructs were inserted into the vector pcDNA3.1- (Invitrogen).

Cell surface protein biotinylation

24 hours after transfection cells expressing the RGMa protein isoforms were washed with ice cold PBS pH 8.0 and incubated with 0.5 mg/ml EZ-Link Sulfo-NHS-LC-Biotin (Thermo Scientific) diluted in PBS pH 8.0 on ice for one hour. Cells were washed twice with ice cold PBS pH8.0 and once with ice cold 25 mM Tris-HCl, pH 8.0, lysed with 200 µl RIPA buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 0.5% sodium deoxycholate, 1% Nonidet P-40, 0.1% SDS) containing proteinase inhibitor cocktail Complete (Roche Diagnostics GmbH, Mannheim, Germany) and cell lysates were cleared by centrifugation. Cleared cell lysates with a total protein amount of 150 µg adjusted to a concentration of 1µg/µl were incubated under rotation with 100µl of a 50:50 slurry of NeutrAvidin^{TM} agarose resins (resuspended in RIPA buffer) over night. The supernatants were collected and the resins were washed three times with RIPA buffer, before supplemented with protein sample buffer (Roti-Load 1, Carl Roth, Karlsruhe, Germany) and subjected to SDS-PAGE and Western blot ananlysis.

Cytosol-ER fractionation

Cytosolic and microsomal fractions of transfected HEK 293T cells were obtained using the Qproteome Mitochondria Isolation Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. 20 µg of protein of each fraction was supplemented with protein sample buffer (Roti-Load 1, Carl Roth, Karlsruhe, Germany) and separated by SDS-PAGE and analysed by Western blot.

Supplementary Data

Supplemental Table S1 Predicted targeting capacities of the N-domain and the full length SP of the RGMa protein isoforms 1 and 2. The isolated N-domains and the full-length SP sequences were analyzed *in silico* regarding their targeting capacity using the software *TargetP v1.1*[1] and *SignalP v3.0*[2-4], respectively. The score of *TargetP v1.1* is an artificial neural network output, the higher the more probable the sequence is a mitochondrial targeting peptide. The score of *SignalP v3.0* indicates the probability for the sequence to be a secretory pathway signal peptide.

	iso1-N-RGMa	iso2-N-RGMa	iso1-SP-RGMa	iso2-SP-RGMa
TargetP v1.1	0.6	0.1	0.7	0.4
SignalP v3.0	0.0	0.0	0.9	1.0

Supplemental Figure S1

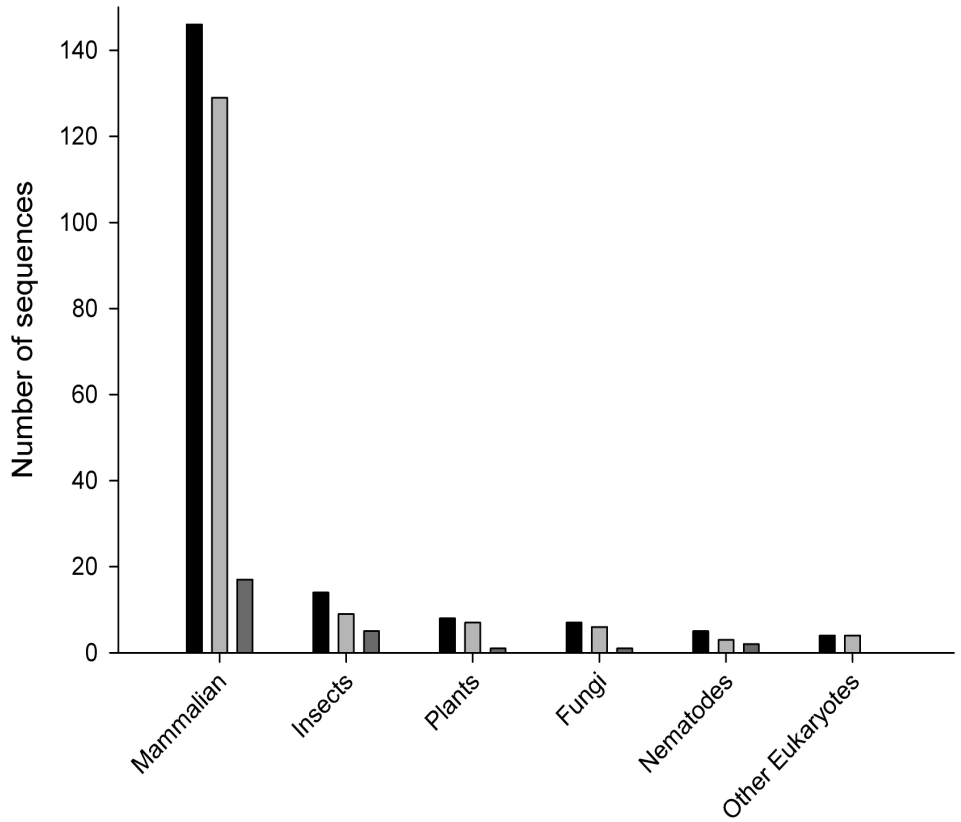

Fig. S1 Species origin of eukaryotic long signal peptides with more than 40 residues and evidence for the protein to exist at protein level (*N*=136) in the UniProtKB Database.[5] Black: total number of signal sequences. Light gray: signal sequences predicted to be NtraC organized. Dark gray: signal sequences not predicted to be NtraC organized.

Supplemental Figure S2

A

```
CLUSTAL 2.0.8 multiple sequence alignment

SeqA Name                        Len(aa)  SeqB Name                        Len(aa)  Score
================================================================================================
1    UniProtKB|Q96B86|RGMA_HUMAN   450      2    UniProtKB|Q6NW40|RGMB_HUMAN   437      47
1    UniProtKB|Q96B86|RGMA_HUMAN   450      3    UniProtKB|Q6ZVN8|RGMC_HUMAN   426      47
2    UniProtKB|Q6NW40|RGMB_HUMAN   437      3    UniProtKB|Q6ZVN8|RGMC_HUMAN   426      41
================================================================================================


UniProtKB|Q96B86|RGMA_HUMAN      MQPPRERLVVTGRAGWMGMGRGAGRS--ALGFWPTLAFLLCSFPA ----A 44
UniProtKB|Q6NW40|RGMB_HUMAN      MGLRAAPSSAAAAAAEVEQRRRPGLCPPPLELLLLLLFSLGLLHA▾GDCQQ 50
UniProtKB|Q6ZVN8|RGMC_HUMAN      ----------MGEPGQSPSPRSSHGSPPTLSTLTLLLLLCCHAHS▾----- 35
                                           *       *     *

UniProtKB|Q96B86|RGMA_HUMAN      TSP▾CKILKCNSEFWSATSGSHAPASD---------------DTPEFCAAL 79
UniProtKB|Q6NW40|RGMB_HUMAN      PAQ CRIQKCTTDFVSLTSHLNSAVDG---------------FDSEFCKAL 85
UniProtKB|Q6ZVN8|RGMC_HUMAN      --Q CKILRCNAEYVSSTLSLRGGGSSGALRGGGGGGRGGGVGSGGLCRAL 83
                                    * *  *     * *                          * **
```

B

```
CLUSTAL 2.0.8 multiple sequence alignment

SeqA Name                        Len(aa)  SeqB Name                        Len(aa)  Score
================================================================================================
1    UniProtKB|Q96PD2|DCBD2_HUMAN  775      2    UniProtKB|Q91ZV3|DCBD2_MOUSE 769      84
1    UniProtKB|Q96PD2|DCBD2_HUMAN  775      3    UniProtKB|Q91ZV2|DCBD2_RAT   769      85
1    UniProtKB|Q96PD2|DCBD2_HUMAN  775      4    gi|114588144|P.troglodytes   775      99
1    UniProtKB|Q96PD2|DCBD2_HUMAN  775      5    gi|119879152|B.taurus        770      91
2    UniProtKB|Q91ZV3|DCBD2_MOUSE  769      3    UniProtKB|Q91ZV2|DCBD2_RAT   769      92
2    UniProtKB|Q91ZV3|DCBD2_MOUSE  769      4    gi|114588144|P.troglodytes   775      84
2    UniProtKB|Q91ZV3|DCBD2_MOUSE  769      5    gi|119879152|B.taurus        770      83
3    UniProtKB|Q91ZV2|DCBD2_RAT    769      4    gi|114588144|P.troglodytes   775      85
3    UniProtKB|Q91ZV2|DCBD2_RAT    769      5    gi|119879152|B.taurus        770      83
4    gi|114588144|P.troglodytes    775      5    gi|119879152|B.taurus        770      90
================================================================================================

UniProtKB|Q96PD2|DCBD2_HUMAN     MASRAVVRARRCPQCPQVRAAAAAPAWAALPLSRSLPP----CSNSSSFS   46
UniProtKB|Q91ZV3|DCBD2_MOUSE     MASRAPLRAARSPQGPGGPAAPAATGRAALPSAGCCPLP--PGRNSSSRP   48
UniProtKB|Q91ZV2|DCBD2_RAT       MASRAPLRAARSPQDPGGRAAPAATGRAPLPSAGWCPLP--PGRNSSSRP   48
XP_001141399|P.troglodytes       MASRAVVRARRCPQCPQVRAAAAAPAWAALPLSRSLPP----CSNSSSFS   46
XP_602937|B.taurus               MASRAVVRAGHSPQRFLVRAAVAAPARAAFPLSRSYPLPFRSNSSSTSFP   50
                                 *****  **   **     ** **   *  *     *       * *

UniProtKB|Q96PD2|DCBD2_HUMAN     MPLFLLLLLVLLLLLEDAGA▾QQGDGCGHTVLGPESGTLTSINYPQTYPNS   96
UniProtKB|Q91ZV3|DCBD2_MOUSE     R-----LLLLLLLLLQDAGG▾QQGDGCGHTVLGPESGTLTSINYPHTYPNS   93
UniProtKB|Q91ZV2|DCBD2_RAT       R-----LLLLLLLLLPDAGA▾QKGDGCGHTVLGPESGTLTSINYPHTYPNS   93
XP_001141399|P.troglodytes       MPLFLLLLLVLLLLLEDAGA▾QQGDGCGHTVLGPESGTLTSINYPQTYPNS   96
XP_602937|B.taurus               RPLFLLLLLILLLLLEDAGA▾QQGDGCGHTVLGPESGTLTSINYPHTYPNS  100
                                  *** ***** *** * *** ********************* *****
```

Fig. S2 The NtraC organization of RGM's and DCBLD2's signal peptide. (A) Multiple sequence alignment of the N-terminus of human RGMa (Q96B86), RGMb (Q6NW40) and RGMc (Q6ZVN8) using CLUSTAL 2.0.8.[6] Arrowheads mark by *SignalP v3.0*[2-4] predicted signal peptide cleavage sites. Transition areas according to the NtraC model are underlined. Asterisks mark residues conserved in all aligned sequences. (B) Multiple sequence alignment of the N-terminus of DCBLD2s from H. sapiens (Q96PD2), M. musculus (Q91ZV3), R. norvegicus (Q91ZV2), P. troglodytes (XP_001141399) and B. taurus (XP_602937) using CLUSTAL 2.0.8.[6] Arrowheads mark by *SignalP v3.0*[2-4] predicted signal peptide cleavage sites. Transition areas according to the NtraC model are underlined. Asterisks mark residues conserved in all aligned sequences.

Supplemental Figure S3

A

B

construct	signal sequence
SP-RGMa	MQPPRERLVVTGRAGWMGMGRGAGRSALGFWPTLAFLLCSFPAATSP_CKIL
N-RGMa	MQPPRERLVVTGRAGWMGMGRGAGRS
C-RGMa	MALGFWPTLAFLLCSFPAATSP_CKIL
SP-DCBLD2	MASRAVVRARRCPQCPQVRAAAAAPAWAALPLSRSLPPCSNSSSFSMPLFLLLLLVLLLLLEDAGA_QQG
N-DCBLD2	MASRAVVRARRCPQCPQVRAAAAAPAWAALPLSRSLPPCSNS
C-DCBLD2	MSSFSMPLFLLLLLVLLLLLEDAGA_QQG

C

construct	signal sequence
iso1-N-RGMa	MGGLGPRRAGTSRERLVVTGRAGWMGMGRGAGRS
iso2-N-RGMa	MGMGRGAGRS
iso1-SP-RGMa	MGGLGPRRAGTSRERLVVTGRAGWMGMGRGAGRSALGFWPTLAFLLCSFPAATSP_CKIL
iso2-SP-RGMa	MGMGRGAGRSALGFWPTLAFLLCSFPAATSP_CKIL

Fig. S3 Blueprint of the RGMa and DCBLD2 SEAP constructs. (A) Assembly of the SEAP fusion protein. Constructs were generated by fusion of the signal sequence of interest (see B and C) to SEAP, lacking the endogenous signal sequence ($SEAP^{\Delta SP}$) and an *EcoR*I restriction site as a linker. All fusion proteins are labeled with a myc-tag at the C-terminus. (B) Overview of the N-termini of the generated SEAP constructs with the assigned signal sequences. SP: native signal sequence, N: N-domain of the signal sequence, C: C-domain of the signal sequence. Underscore: predicted signal peptide cleavage site. RGMa sequences refer to RGMa protein isoform 3 (NCBI accession NP_064596.2). (C) Overview of the N-termini of the generated SEAP constructs with the assigned signal sequences. SP: native signal sequence, N: N-domain of the signal sequence, Underscore: predicted signal peptide cleavage site. iso1-(N/SP)-RGMa and iso2-(N/SP)-RGMa sequences refer to RGMa protein isoform 1 (NCBI accession NP_001159755.1) and isoform 2 (NCBI accession NP_001159758.1), respectively.

Supplemental Figure S4

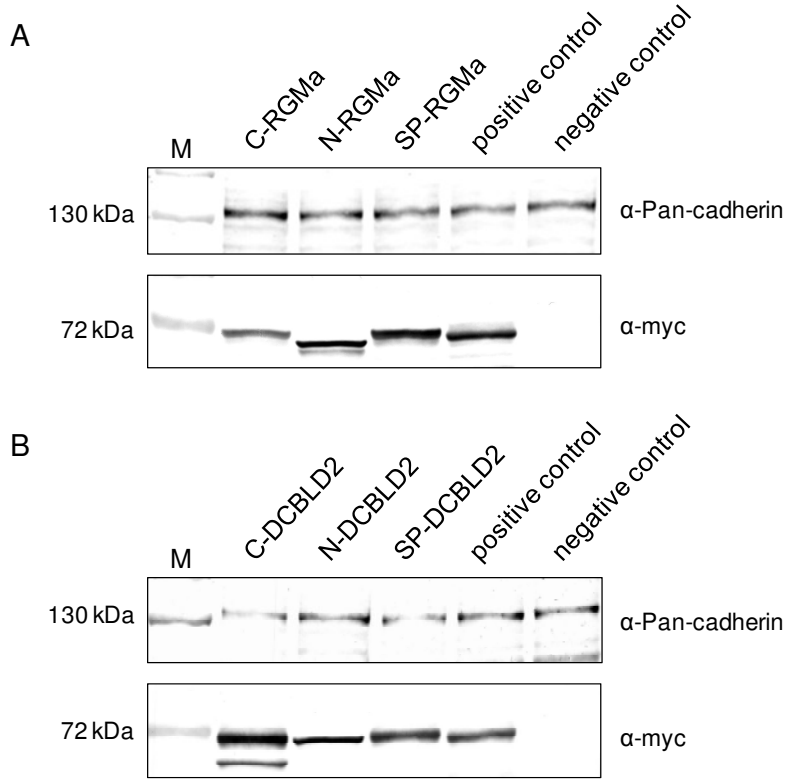

Fig. S4 Expression of the SEAP constructs in HEK 293T cells. (A) Detection of RGMa and (B) DCBLD2 SEAP fusion proteins by Western blot analysis. An amount of 20 µg of protein per sample was separated by SDS-PAGE. Fusion proteins were detected with α-myc antibody. Positive control: Lysate of cells transfected with the construct coding for the unmodified SEAP protein tagged with myc. Negative control: Lysate of cells transfected with the empty vector (pcDNA3.1-). As a loading control endogenous protein was detected with α-Pan-cadherin antibody. M: protein ladder.

Supplemental Figure S5

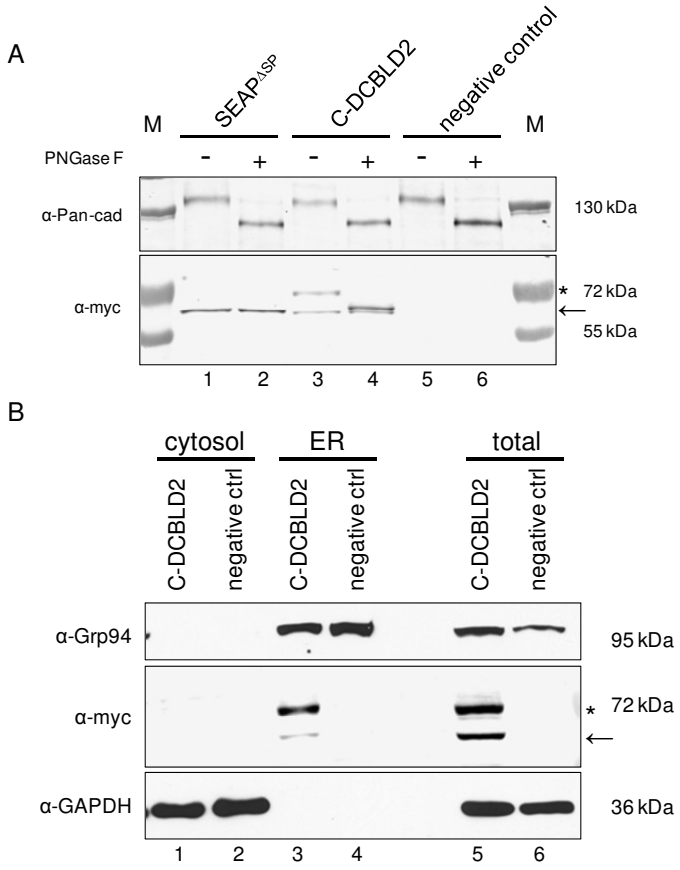

Fig. S5 Characterisation of the lower migrating protein band of C-DCBLD-2. (A) The PNGase F digest of C-DCBLD-2 protein expressed in HEK 293T (cf. Fig. 3B, lanes 3 and 4) was replicated and analysed by SDS-PAGE and Western blot using an alternative gel system. Protein samples of HEK 293T cells expressing SEAP$^{\Delta SP}$, C-DCBLD-2, or the empty vector (negative control) were subjected to PNGase F digestion (+) removing putative *N*-glycans or were left untreated (-) and 10µg of total protein per sample was analyzed by Western blot using α-myc antibody. For verification of PNGase F digestion cadherins were detected with α-Pan-cadherin antibody. The lower migrating band of C-DCBLD-2 is located at the size of SEAP$^{\Delta SP}$ lacking any signal sequence (lane 2 and 3, arrow) suggesting an N-terminal peptide cleavage by signal peptidase an therefore ER-lumenal localisation. Contrarily, this protein species does not shift after PNGase F treatement suggesting to be not N-glycosylated. The upper band of C-DCBLD-2 shows PNGase F sensitivity (lane 3 and 4, asterisk and arrow). Notably, the shift of the upper band does not reach the lower migrating band generating a doublet (lane 4, arrow) indicating that different signal peptidase cleavage sites may be used for the two protein species. M: protein ladder. (B) The cytosol-ER fractionation of HEK 293T cells expressing C-DCBLD-2 shows that the lower migrating protein species as the higher migrating species are located in the ER fraction (lane 3, asterisk and arrow). Purity of ER and cytosol fractions was determined by detecting Grp94 and GAPDH as ER and cytosolic markers, respectively. negative ctrl: cells transfected with the empty vector (pcDNA3.1-).

Supplemental Figure S6

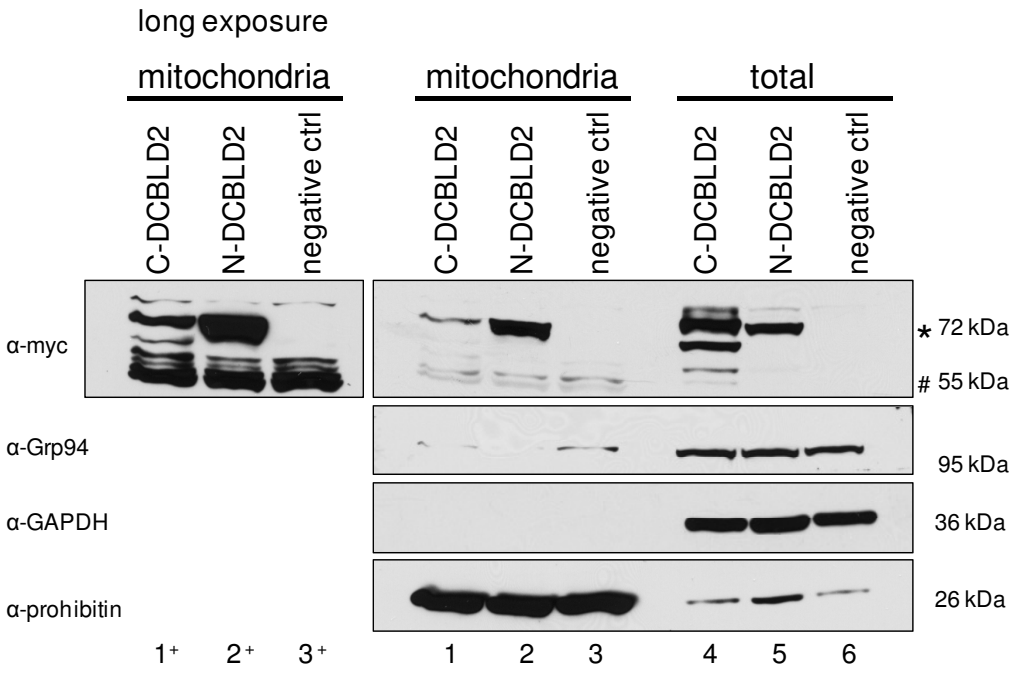

Fig. S6 Clarification of the unspecific signal of the anti-myc antibody in the mitochodrial fraction (cf. Fig. 5, hash). The mitochonrial fractions (mitochondria) of C-DCBLD-2 and N-DCBLD-2 expressing cells were complemented by the mitochondrial fraction of HEK 293T cells transfected with the empty vector (negative ctrl). SDS-PAGE in an alternative gel-system with larger separation gels and Western blot analysis replicated the signals of the anti-myc antibody revealing that the signal about 55 kDa is composed of several bands which are also found in the negative control (long exposure, lane 1^+, 2^+, 3^+, hash) indicating that this signal is a cross reaction of the anti-myc antibody with proteins of the mitochondrial fraction. A shorter exposure weakened this signal (lane 1, 2, 3, hash). Grp94 (ER marker protein) and GAPDH (cytosolic protein) were detected to depict the purity of mitochondrial fraction. Prohibitin (mitochondrial protein) was visualised as a mitochondrial marker using mouse monoclonal anti-prohibitin antibody (1.0 µg/ml; medac, Wedel, Germany).

Supplemental Figure S7

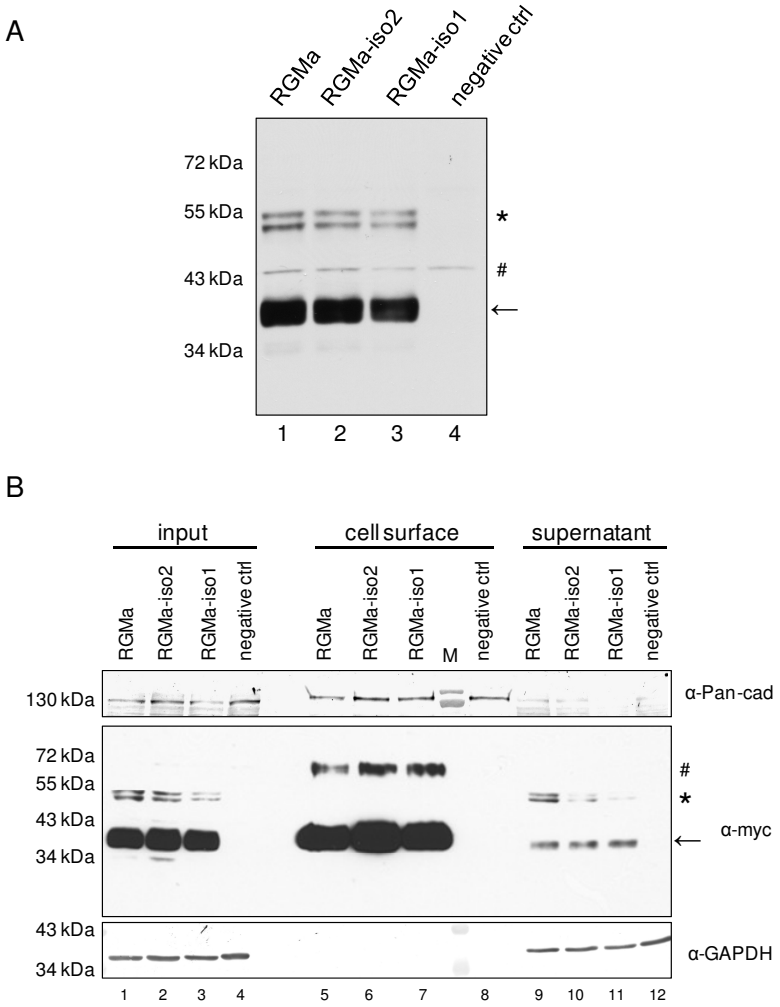

Fig. S7 Expression of RGMa protein isoform 1, 2, and 3 in HEK 293T cells. HEK 293T cells were transfected with constructs coding for the RGMa isoforms (isoform 1: RGMa-iso1; isoform 2: RGMa-iso2; isoform3: RGMa) or with the empty vector (negative ctrl). (A) Western blot analysis was performed using a rabbit polyclonal α-RGMa antibody. The pattern in the Western blot revealing a mature protein species (arrow) and the precursor asterisk) is comparable to the pattern observed for the expression of chicken RGMa in HEK 293 cells.[7] Hash: unspecific signal of the antibody. (B) Cell surface proteins of RGMa protein isoforms expressing HEK 293T cells were probed with biotin, pulled down with NeutrAvidin beads and analysed by SDS-PAGE and Western blot (cell surface). 10% of the pull-down input (input) and the supernatant (supernatant) were also subjected to the Western blot. The mature protein isoforms (arrow) are present in the cell surface samples, but the precursor proteins (asterisk) are not. A not further definable protein band of about 70 kDa is detectable in the cell surface samples (hash). It might be derived of RGMa protein, as it is only present in the samples of RGMa expressing cells, but being of too low amount for detection in the whole cell lysate. Cadherins were visualised with α-Pan-cadherin antibody to show comparable pull-down efficiencies among the cell surface samples. GAPDH, an intracellular protein, was detected to check the purity of the cell surface samples. M: protein ladder.

Supplemental Figure S8

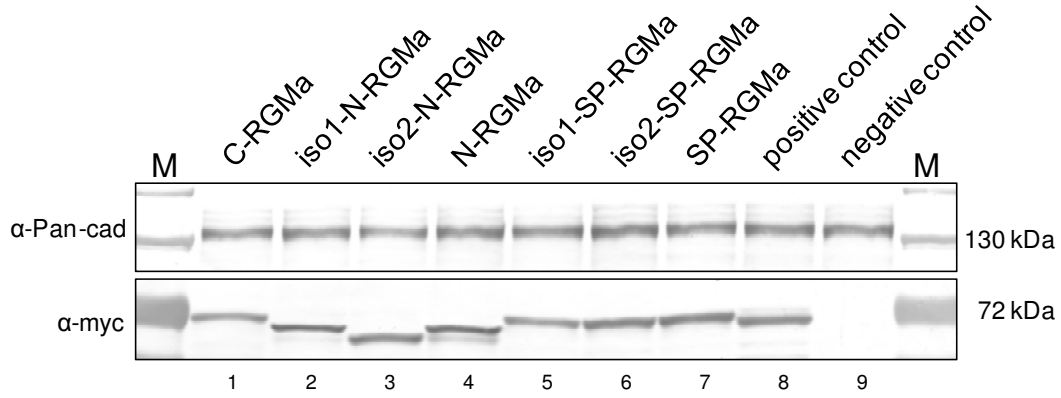

Fig. S8 Expression of the RGMa SEAP constructs exhibiting the signal sequences of the RGMa protein isoforms in HEK 293T cells. For detailed illustration of the signal sequences compare Supplemental Fig. S3. Detection of RGMa SEAP fusion proteins by Western blot analysis. An amount of 20 µg of protein per sample was separated by SDS-PAGE. Fusion proteins were detected with α-myc antibody. Positive control: Lysate of cells transfected with the construct coding for the unmodified SEAP protein tagged with myc. Negative control: Lysate of cells transfected with the empty vector (pcDNA3.1-). As a loading control endogenous protein was detected with α-Pan-cadherin antibody. M: protein ladder.

Supplemental Figure S9

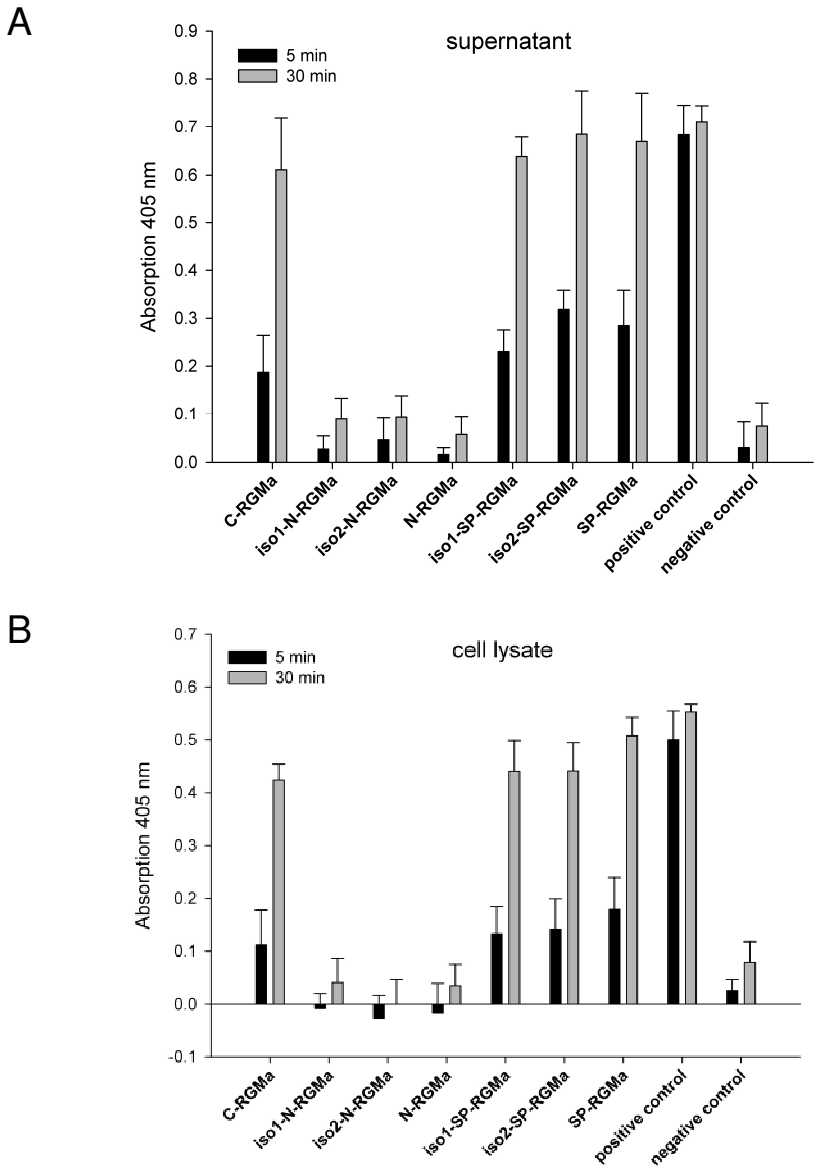

Fig. S9 SEAP activity of the RGMa SEAP fusion proteins comprising the signal sequences of the RGMa protein isoforms. The enzymatic activity was determined in the supernatant and in the lysate of transfected HEK 293T by incubation of the (A) supernatant or (B) cell lysate with the substrate *p*-nitrophenyl-phosphate. The absorption of the product (nitrophenol) was measured at the wavelength of 405 nm after 5 min and 30 min. Error bars indicate S.D. ($N \geq 4$). The activities of the corresponding constructs carrying the SPs of RGMa isoforms (iso1-SP-RGMa, iso2-SP-RGMa, and (iso3-)SP-RGMa) or the N-domains (iso1-N-RGMa, iso2-N-RGMa, and (iso3-)N-RGMa), respectively, show no prominent difference among each other. SP: native signal sequence, N: N-domain of the signal sequence, C: C-domain of the signal sequence. iso1, iso2: isoform 1 and isoform 2, respectively (for detailed illustration of the signal sequences compare Supplemental Fig. S3). Positive control: Cells transfected with the construct coding for the unmodified SEAP protein tagged with myc. Negative control: Cells transfected with the empty vector (pcDNA3.1-).

Supplemental Figure S10

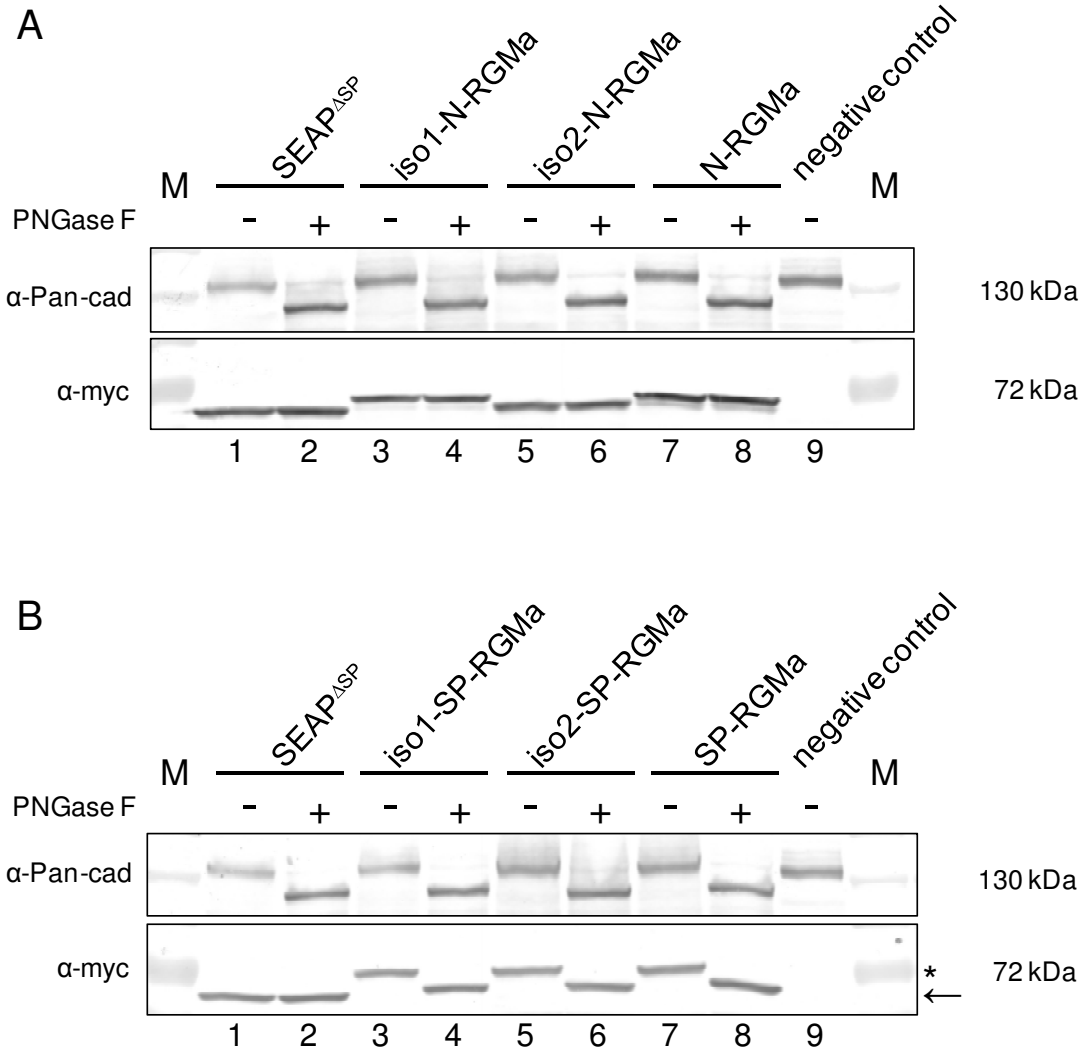

Fig. S10 Western blot of PNGase F treated RGMa SEAP fusion proteins. (A) SEAP fusion proteins exhibiting the N-domains (A) and the full length SPs (B) of RGMa isoform 1 (iso1-N/SP-RGMa), isoform 2 (iso2-N/SP-RGMa), and isoform 3 (N/SP-RGMa) were tested for PNGase F sensitivity as a marker for protein translocation to the ER. Protein lysates of transfected HEK 293T cells were subjected to peptide-*N*-glycosidase F (PNGase F) digestion (+) removing putative *N*-glycans or were left untreated (-) and analyzed by Western blot using α-myc antibody. For verification of PNGase F digestion cadherins were detected with α-Pan-cadherin antibody. SEAP$^{\Delta SP}$: SEAP lacking any signal peptide. Negative control: Lysate of cells transfected with the empty vector (pcDNA3.1-). Asterisks: *N*-glycosylated proteins. Arrows: shifted PNGase F deglycosylated proteins. M: protein ladder.

Supplemental Figure S11

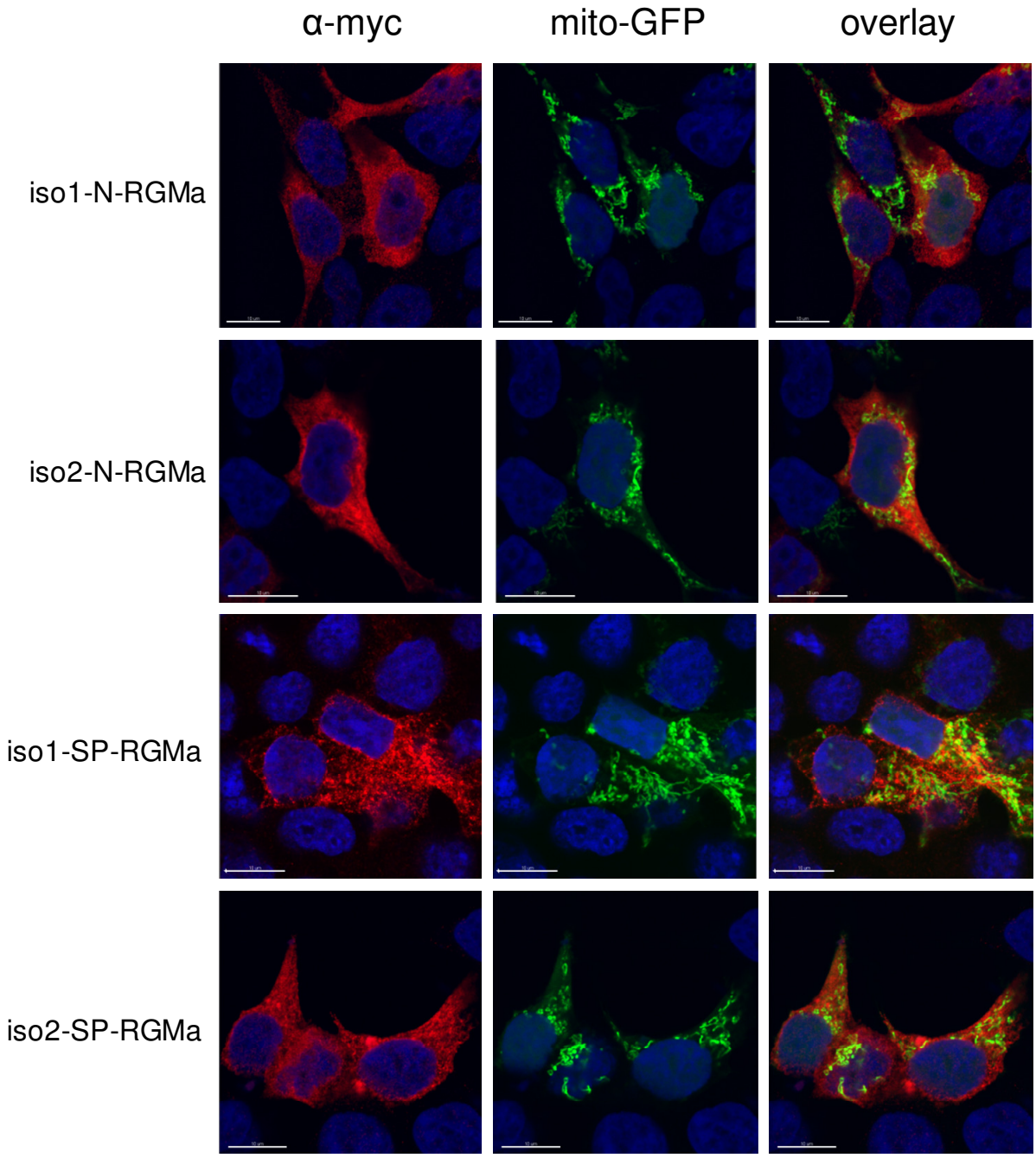

Fig. S11 Localization of the RGMa SEAP fusion proteins carrying the N-domains (iso1/2-N-RGMa) and the full length SPs (iso1/2-SP-RGMa) of isoform 1 and 2, respectively. For detailed illustration of the signal sequences compare Supplemental Fig. S3C. HEK 293T cells were cotransfected with the constructs and mito-GFP plasmid. SEAP fusion proteins were visualized using α-myc as primary antibody and Alexa 594 conjugated secondary antibody (red, left lane), mitochondria were stained using mito-GFP (green, center lane). Nuclei were stained with DAPI. Overlay of the channels is shown in the right lane. Scale bar: 10 µm.

References

1 O. Emanuelsson, H. Nielsen, S. Brunak and G. von Heijne, *J Mol Biol*, 2000, **300**, 1005-1016.
2 J. D. Bendtsen, H. Nielsen, G. von Heijne and S. Brunak, *J Mol Biol*, 2004, **340**, 783-795.
3 H. Nielsen, J. Engelbrecht, S. Brunak and G. von Heijne, *Protein Eng*, 1997, **10**, 1-6.
4 H. Nielsen and A. Krogh, *Proc Int Conf Intell Syst Mol Biol*, 1998, **6**, 122-130.
5 C. H. Wu, R. Apweiler, A. Bairoch, D. A. Natale, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, R. Mazumder, C. O'Donovan, N. Redaschi, et al., *Nucleic Acids Res*, 2006, **34**, D187-191.
6 J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res*, 1994, **22**, 4673-4680.
7 P. P. Monnier, A. Sierra, P. Macchi, L. Deitinghoff, J. S. Andersen, M. Mann, M. Flad, M. R. Hornberger, B. Stahl, F. Bonhoeffer and B. K. Mueller, *Nature*, 2002, **419**, 392-395.